

Feedback strategies for visual search in airframe structural inspection

Anand K. Gramopadhye^{a,*}, Colin G. Drury^b, Joseph Sharit^b

^a *Department of Industrial Engineering, Clemson University, Clemson, SC, USA*

^b *Department of Industrial Engineering, State University of New York, Buffalo, NY, USA*

Received 28 August 1995; revised 17 January 1996

Abstract

Feedback of information has consistently shown positive results in human inspection, provided it is given in a timely and appropriate manner. Feedback serves as the basis of most training schemes; traditionally this has been performance feedback. Other forms of feedback which provide strategy information rather than performance information may have a role in improving inspection. This study compared performance feedback and cognitive feedback in a realistic simulation of an aircraft structural inspection task. Performance (time, errors) feedback showed the greatest improvements in performance measures. Cognitive feedback enhanced efficiency measures of search strategy. When cognitive feedback consisted of visual representations of the path and the coverage of the search sequence, subjects also were able to use this task information to improve their search performance.

Relevance to industry

The results of this study have direct implications on developing training strategies for improving industrial inspection performance. The results can be used to design superior training programs to improve airframe inspection and thus aviation safety. The results also can be extended to complex inspection tasks in other industries (printed circuit board inspection in the electronic industry, cloth inspection in the textile industry, inspection of food products in the agricultural industry and in-process inspection in manufacturing industry).

Keywords: Feedback; Visual research; Inspection

1. Introduction

Continuing airworthiness of aircraft structures is largely based upon the visual detection of small defects by trained inspection personnel (Shepherd et

al., 1991). Inspection can be modeled as a two-stage process with (1) a search for potential faults leading to (2) a decision as to whether or not the fault meets pre-specified reporting criteria (Drury, 1990). If defects are to be detected reliably, the search task, in this case visual search, is thus the first aspect of human performance required. Visual search of large visual fields has a long history of study and model-

* Corresponding author.

ing to guide current applications, e.g., Lamar (1960), Morris and Horne (1960), National Academy of Sciences (1973), special issue of *Human Factors* (Human Factors Society, 1979), Brogan (1990) and Brogan et al. (1993). This research examines one aspect of human/system fit in search particularly appropriate to aircraft structural inspection: the training scheme.

Studies in visual search have shown that both speed and accuracy can be improved with training (Parkes and Rennocks, 1971; Bloomfield, 1975). Practice on search tasks has shown encouraging results in improving visual search performance (Baker et al., 1960; Chaney and Teel, 1967; Drury and Clement, 1978; Parker and Perry, 1972). More recently, Drury and Kleiner (1993) showed how search training could be incorporated into a successful training program for aircraft bearings. Thus in both inspection and its search component we know that training can be effective.

However, there are still unresolved issues in understanding the role of specific training interventions (e.g., Gramopadhye et al., 1993). For example, models of search (e.g., Morawski et al., 1980; Greening, 1976) have shown that search performance depends on three factors: (1) the time taken for one fixation, (2) the area covered in one fixation (visual lobe), and (3) the strategy adopted in covering the search field with successive visual lobes. How are these affected by the different training interventions? Can we tailor interventions to specific requirements of the task? This paper examines one well-known training intervention, feedback, to determine how different types of feedback affect search strategy and hence performance.

Traditionally, feedback provided to the inspectors has been performance feedback which consists of feedback on search times, search errors (faults not detected) and decision errors (Chaney and Teel, 1967; Cockrell and Sadacca, 1971; Drury and Addison, 1973; Embrey, 1975; Czaja and Drury, 1981; Micalizzi and Goldberg, 1989). Investigators have reported improved performance (reduced search times, reduced search errors and decision errors) by providing performance feedback. Weiner (1975) reviewed feedback in training for inspection and vigilance, and found it to be universally beneficial. However, it is possible to give feedback which provides task infor-

mation on the process, or strategy, by which the inspector achieved these results. This kind of feedback is referred to as 'cognitive feedback'. Although cognitive feedback has not been applied to inspection situations per se, it has been shown to have a beneficial effect for non-inspection situations. Todd and Hammond (1965) used a multiple cue probability learning task and found that providing cognitive feedback led to significantly higher achievements than objective performance feedback. Hammond (1971) found that cognitive feedback (feedback about the task characteristics) led to rapid learning of a judgement task. Furthermore, Hammond et al. (1973) found that outcome feedback led to deterioration in performance in relation to cognitive feedback. In a study conducted by Deane et al. (1972), cognitive feedback was found to result in a better understanding of the task characteristics and in improved control by humans over the execution of their knowledge (i.e., better cognitive control). Lindell (1976) found evidence to suggest the superiority of cognitive feedback over outcome feedback for a simple learning task. Doherty and Balzer (1988) also have listed studies that prove the superiority of the relational information provided by cognitive feedback over outcome feedback, at least for the multiple cue probability learning task, where outcome feedback may prove unusable as task complexity increases. In a more recent review of a variety of decision studies (Balzer et al., 1989), both in realistic and abstract tasks, cognitive feedback showed consistent success in providing individuals insight into their own policies and strategies.

Decision studies have used many different cognitive feedback modes, for example auditory, graphical and statistical (Balzer et al., 1989). All have had some success in giving individuals insight into their own policies or strategies. For visual search, auditory feedback would seem inappropriate as input is primarily visual and spatial, with output typically a motor action. Within these modes of 'cognitive' feedback, there are different ways in which information can be presented to the inspector. Cognitive feedback could consist of: (1) statistical feedback which provides statistical information (e.g., the percentage search area covered, number of fixations used to inspect the area, mean inter-fixation distance, percentage overlap) and (2) graphical feedback (vis-

ual feedback of the scan pattern). However, in the visual search context, cognitive feedback may or may not help the inspector to improve search strategy – evaluation is needed. The literature is silent on which is the better form of cognitive feedback, or whether either is better than traditional performance feedback.

In a temporally-extended visual search task, which proceeds by successive fixations, strategy has come to mean the spatial, or at least the logical pattern of successive fixation locations. Thus, a random search strategy (e.g., Krendel and Wodinski, 1960) is one in which each fixation position is chosen independently of previous fixations. In contrast, a systematic strategy is where each fixation position is chosen only from those positions as yet unsampled (e.g., Williams, 1966). These two extreme strategies represent end points on a continuum of the searcher's memory for previous position (Arani et al., 1984). As systematic search strategy is more efficient (Tsao et al., 1979; Karwan et al., 1995), cognitive feedback on the visual search strategy should encourage its adoption, and hence improve search performance.

To study the effect of different feedback training strategies on visual search performance, a practical visual inspection task was chosen – the inspection of airframe structures for visual defects. The importance of visual inspection in aircraft inspection is that it accounts for almost 90% of aircraft inspection (Drury et al., 1990). Since the time an aircraft spends in maintenance represents a large loss in revenue, the inspection system must combine effectiveness with efficiency. Training is an important way in which airlines seek to improve human reliability in aircraft maintenance and inspection. Training for inspection tends to be largely on the job, which may not be the most effective or efficient method of instruction (Shepherd et al., 1991). As part of a larger program to improve aircraft inspection performance, the current experiment was undertaken to explore different ways of enhancing search strategy and hence performance. Cognitive feedback, presented as either statistical or graphical information, was compared to both performance (outcome) feedback and a control (no feedback) condition, using a computer simulation of an aircraft structural visual inspection task. The computer simulation's physical attributes and functional characteristics were based on task analyses of

inspection tasks (Drury et al., 1990; Gramopadhye et al., in press). The inspection task required the subject to successively move a viewer (representing a flashlight beam) around each screen of the task in order to detect airframe structural defects. Movements of the viewer were used as the basis for cognitive feedback of search strategy.

2. Experiment

2.1. Methods

2.1.1. Subjects

The subjects were 24 graduate and undergraduate students of the State University of New York at Buffalo, in the age group of 20–30 years. Gallwey and Drury (1986) have shown that minimal differences exist between inspector and student subjects. Subjects were tested for 20/20 vision (corrected if required) and paid \$5.00/hr for their time. Six subjects were randomly assigned to four different groups: (1) control, (2) statistical process measures, (3) graphical (visual), and (4) performance measures. The distinctions between these groups are discussed in more detail below.

Control group: Subjects assigned to this group underwent only practice between trial 1 and trial 2 of the criterion task.

Statistical process measures group: Subjects in this group were provided with statistical feedback on process measures during the training session between trials 1 and 2.

Graphical (visual) group: Subjects in this group were provided with graphical feedback of the search pattern adopted during the training session between trial 1 and trial 2.

Performance measures group: Subjects assigned to this group received feedback on performance measures (times, errors) during the training session.

2.1.2. Stimulus material

The task was a simulated visual inspection task of airframe inspection implemented on a Sun SPARC work station, a high performance, high resolution system (1152 × 900 pixels, 99.2 dots/inch). The input devices were a standard keyboard and an optical, three button mouse. The task consisted of in-

specting a part of the aircraft fuselage, simulating aluminum alloy skin held together by rivets. The inspector searched the rivets and the area around the rivets for four types of faults. Two types of faults could occur on rivets themselves: (1) rivet cracks – indicated by a longitudinal hairline crack along the rivet edge, and (2) loose rivets – represented by accumulation of rings of dirt along the circumference of the rivet edge. Two types of faults could occur in the area around the rivets: (1) dents – shown as rippled metal, and (2) corrosion – represented as a collection of small grey and white globules on the metal skin.

2.1.3. Visual search task

The criterion task was an unpaced visual search task, using the airframe inspection simulator, where subjects searched for one example of one of the four fault types on each area. Areas contained either zero or one fault. The entire inspection task was divided into a series of search areas, with a search area defined as that portion of the task shown on one screen. As each part of the aircraft fuselage (one search area) was presented to the subjects, their task was to locate any fault in the search area. They indicated that they had located the fault by clicking the left mouse button on the fault.

To simulate the use of local lighting, such as a flashlight, not all of the area was available for inspection at any one time. A movable window, i.e. the field of view or 'viewer' as it was explained to the subjects, could be moved around the area using the mouse, exposing whatever was within the viewer's field of view. The whole area subtended a visual angle of 22° by 22° while the viewer subtended an angle of 6° by 5° at the eye position 500 mm from the screen. The subjects inspected any region of the area by clicking the left mouse button on that region, upon which action the center of the viewer would move to the point clicked, thereby illuminating a rectangular area around the center point. If the subjects detected a fault within the viewer area, they responded by clicking the left mouse button on the fault, and classified it by entering the decision code corresponding to the fault detected. Once the fault was located and classified, subjects moved to the next area by clicking on 'NEXT' on the screen. Each visual search trial consisted of 75 randomly ordered

search areas, 60 of which contained single faults (15 of each of the 4 types) while 15 areas had no fault.

2.1.4. Training procedure

All the subjects were administered two tests measuring cognitive styles pertinent to inspection (the Matching Familiar Figures Test; Schwabish and Drury (1984) and Embedded Figures Test; Gallwey (1986); Drury and Wang (1986)) and were provided with a brief description of the experiment. Before the start of the first trial, subjects were shown the entire area to be inspected and provided with a graphical and verbal description of all the four faults. Subjects were allowed to experiment with the computer set-up until they were comfortable using it. Following this step, the subjects were provided with a demonstration program to help familiarize them with the task. The demonstration consisted of five areas. Four of the areas contained a single fault of one type, and one area contained no fault. The subjects searched and classified the faults as they would in the trial session.

As noted earlier, subjects were randomly assigned to four different experimental groups. Subjects in all four groups performed the visual search task (trial 1). Following the completion of trial 1, subjects received three practice visual search tasks similar to trial 1, the only difference being that the tasks consisted of 25 search areas each, 20 areas with a single fault type and five areas with no faults, randomly ordered.

Subjects in the Control group did not receive any feedback on the practice trials. The Statistical Process group received statistical feedback on process measures after inspecting each area in all three practice trials. The statistical feedback was divided into two parts: (1) immediate feedback for the previous area inspected; and (2) cumulative feedback over all inspected areas. Feedback to the Statistical Process group was provided on four process measures: (1) percentage area covered; (2) number of windows (viewers) used to inspect the area; (3) mean inter-fixation distance for viewer movements; and (4) percentage overlap of successive viewers. The Graphical (visual) group received feedback on the movement of the viewer window after inspecting each area in all three practice trials. For this group, the feedback graphically displayed the pattern adopted by the subjects in moving the viewer. The

viewer was represented by a marker, with the movement of the marker being analogous to the movement of the viewer. The marker had a color that was different from the background shade; the background color changed to the color of the marker as the marker passed over it, indicating that the corresponding areas had been fixated (covered by the viewer). The markers were shaded with the brightest shade representing the last fixation so that the sequence was given by the shade of the color – the darker shade indicated the earlier fixations in sequence while lighter shades indicated more recent fixations. Providing graphical feedback was designed to help the subjects identify areas covered and missed (distribution of fixations), inter-fixation distance, and the pattern adopted. The Performance group was provided feedback on four performance measures after inspecting each area: (1) percentage of rivet and area faults detected; (2) search time (s) for area and rivet faults; (3) stopping time (time at which the subject terminated the search and moved to the next area); and (4) the percentage time spent in each zone, where the entire inspection area was divided into nine zones (3×3). Following the completion of the training, subjects in all the four groups performed trial 2 of the criterion task, which was identical to trial 1.

2.1.5. Experimental design

A 4×2 design was employed consisting of four feedback groups (control, performance, process, visual) with six subjects in each group, and two trials (before, after training); the latter factor was treated as a repeated measure.

2.1.6. Data collection

Data was collected on the process measures and performance measures listed below. Note that these measures were the same ones given to the subjects in the appropriate training groups. The process measures are based on eye movement parameters that contribute to search strategies as defined by Megaw and Richardson (1979). The following sets of measures for field of view (viewer) movements, chosen to be analogous to the eye movement parameters, were thus used to compare the effectiveness and efficiency of different search strategies.

Process measures. The following process measures were used.

1. n = number of viewer fixations,
2. $(X/A) \times 100\%$ = percentage area covered,
3. $1/(n-1) \sum r_i$ = average interfixation distance,
4. $(na - X)/a(n-1) \times 100\%$ = percentage viewer fixation overlap,

where,

- A = total area of inspection screen,
- a = area of the viewer field of view,
- r_i = the length of the i th viewer movement, degrees,
- X = total area covered by the viewer.

Note that $(na - X)$ is the total overlapped area. *Performance measures.* The following performance measures were used:

1. Search time: time (s) from area presentation to fault detection.
2. Stopping time: time (s) at which the subject terminate a search process on an area by deciding that the area does not contain a fault, and the subjects move to the next area (see Gallwey and Drury, 1986, for a discussion on search and stopping times).
3. Fixation time: the average time (s) spent at each viewer location, given by the total time spent in each area divided by the number of viewer fixations.
4. Number of faults detected, recorded separately for each fault type.

Note that subjects in the Performance Feedback condition also were told the total number of faults present so that they could infer if they had missed detecting any fault. Note also that the task was one of search so that false alarms were not expected. Indeed, none were found throughout the experiment.

2.1.7. Verbal protocols

End of session (retrospective) verbal protocols were taken after trial 2 and recorded to help understand the visual search strategy adopted by the subjects and to relate it to the process and performance measures. Following trial 2, subjects performed an inspection task similar to trial 2, but with a shorter duration (consisting of 5 search areas). Subjects were asked to think aloud and verbalize the search pro-

cess. At the conclusion of this session, the subjects were debriefed and were asked to provide a verbal report of the task.

2.2. Results

2.2.1. Analysis of process measures

An analysis of process measures was conducted only on the areas in which the target was not detected (i.e., the trials on which a stopping was recorded) to avoid contamination with searches which were truncated by fault detection. To determine whether the four groups were equivalent before training, ANOVAs of all process measures were performed on trial 1 results only. The only measure for which a significant group effect found was percentage area covered. The control and Graphical (visual) groups covered a smaller percentage of the search area (91%, 92%) than did the Statistical Process and Performance groups (96%, 97%).

The next step of the analysis was directed at determining the effect of training on different process measures relating to visual search performance. Since multiple measures were taken simultaneously, a multivariate analysis of variance (MANOVA) was performed before any univariate ANOVAs. The MANOVA showed a significant Group \times Trial interaction for percentage area covered, number of windows, inter-fixation distance, and percentage overlap, all of which indicated a differential effect of training on the four process measures. The trial effect was highly significant for percentage overlap and number of windows. The group effect was significant for percentage overlap.

Separate univariate ANOVAs conducted on all the process measures showed significant Group \times Trial interactions for percentage area covered, number of windows, inter-fixation distance and percentage overlap. The trial effect was significant for the number of windows and percentage overlap. The group effect was significant for percentage overlap (Table 1). The Statistical Process group showed the largest change (decrease) in the percentage area covered. All the groups (except the Control group) after training showed a decrease in the number of viewer fixations required. The percentage decrease was greatest for the Statistical Process group, followed by the Graphical (visual) group and the Performance

Table 1

F-values obtained from ANOVA (averaged overall faults) conducted on the process measures

		<i>df F</i> -values			
		Percentage area covered	Number of windows	Interfixation distance	Percentage overlap
Group	3	0.85	2.07	0.35	4.05
Trial	1	1.05	15.38 **	0.10	22.58 **
Trial \times Group	3	4.16 *	4.32 *	5.99 *	7.68 **

* $p < 0.05$, ** $p < 0.001$.

group. The inter-fixation distance for all the groups decreased after training except for the Statistical Process group. The percentage overlap after training decreased for all three feedback groups, while that for the Control group increased after training.

2.2.2. Analysis of performance measures

The results of the MANOVA on the mean search time (averaged over all the faults), stopping time and detection accuracy (percentage of faults detected) did not show any significant Group \times Trial interaction or Group effect. The trial effect was significant for mean search time, and for stopping time. For search times and detection accuracy, several levels of analysis are possible because there were four fault types in two groups. ANOVAs across all four fault types, across the two fault type groups (rivet, area), and across each individual fault type were performed to test whether the different fault types were differentially affected by the Group and Trial effects. Only the first analysis (Table 2) is presented here as the others added no separate insights.

An ANOVA on stopping times before and after training showed a significant trial effect, while the

Table 2

F-values obtained from ANOVA (averaged overall faults) conducted on the performance measures

		<i>df F</i> -values			
		Stopping times	Search times	Fixation times	Percentage detected
Group	3	1.25	1.56	3.37 *	2.01
Trial	1	15.03 **	10.63 **	0.10	2.17
Trial \times Group	3	1.90	1.83	0.35	2.01

* $p < 0.05$, ** $p < 0.001$.

Group effect and Group \times Trial interaction effect were not significant. The mean stopping time for process, visual and performance groups decreased significantly after training, but such a trend was not evident for the control group. An ANOVA of fixation time showed a significant Group effect. The performance, control and visual groups showed a decrease in fixation time from trial 1 to trial 2 while the Statistical Process group showed an increase in fixation times.

The ANOVA for mean search time for all four fault types showed a significant Trial effect and a significant Fault type effect. The ANOVA results indicate that the mean search time for fault detection decreased from trial 1 to trial 2. A post hoc test on individual means (Newman-Keuls) showed that dents required significantly less time to detect than all other faults. The search time on corrosion faults decreased with training, with the Performance group showing the largest reduction on both the area faults. However, such a reduction was not observed for the Control group on corrosion faults. A similar reduction in search time was observed on both of the rivet faults for the Statistical Process group, Graphical group and Performance group, with the Performance

group again showing the largest percentage decrease in search time in detecting both of the rivet faults. The Control group showed a minimal increase in search time after training.

ANOVAs conducted on percentage of faults detected for all the four fault types taken together showed a small but significant fault type effect with dents detected most often (94%) and loose rivets detected least often (88%). The percentage of area faults detected before and after training was not significantly different for all the four groups.

2.2.3. Reported visual search strategy

The verbal protocols revealed that subjects perceived themselves to be more systematic after training than before training. During the pre-training trial, the subjects reported adopting a more random search strategy, which could have caused refixating on areas already covered, resulting in a larger overlap and greater number of windows. This trend was observed consistently across all the four groups. Following training, the tendency was for subjects to try and cover the entire area in one scan by adopting a low-overlap systematic strategy (moving the viewer side-by-side horizontally or vertically, covering the

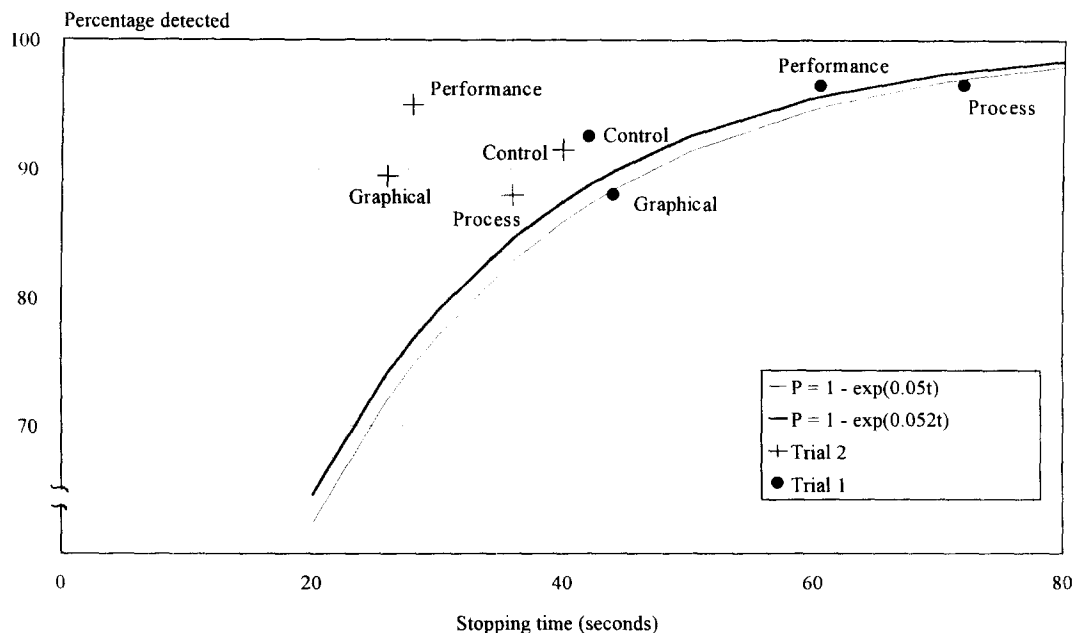


Fig. 1. Speed accuracy tradeoff curve.

entire area in a zig-zag pattern). Subjects went back and searched the area again only if they felt they had not covered a specific area in enough detail.

2.2.4. Speed accuracy tradeoff (SATO)

Because subjects were unconstrained in both speed and accuracy, it is conceivable that separate analyses of these two variables or even the MANOVA would not provide a complete view of the results. For the visual search task, the appropriate SATO is a cumulative negative exponential (e.g., Drury et al., 1980). Fig. 1 shows the mean speed and accuracy results for both the Before and After trials. The four groups in the 'Before' condition were well fitted ($r^2 = 95\%$) by the following cumulative exponential:

$$P(\text{detect}) = 1 - \exp(-0.05t),$$

where, t is the mean stopping time. The Control group has almost the same result in the 'After' condition, while the Statistical Process group increased in speed at the expense of decreased accuracy. If the 'After' results for these two conditions are included in the regression, $r^2 = 97\%$. The value of coefficient changes only by 4% and the exponential equation is:

$$P(\text{detect}) = 1 - \exp(-0.052t).$$

The Graphical (visual) group increased in both speed and accuracy, while the Performance group doubled its speed for almost no change in accuracy. This SATO analysis reinforces the ANOVA conclusions that performance changed mainly for the Graphical and Performance groups.

2.2.5. Correlation of performance and process measures

Correlation analysis was performed on trial 1 process measures (percentage area covered, average number of viewing windows, inter-fixation distance and percentage overlap) and performance measures (search time, stopping time, fixation time and percentage faults detected) for the mean values of all the 24 subjects, to identify the degree of association between the different measures. The intercorrelation matrix of these process and performance measures was subjected to a Factor Analysis using varimax rotation of orthogonal factors. Table 3 shows the two factors which emerged. Factor 1 (with a total vari-

Table 3

Factor loadings obtained from factor analysis (only loadings greater than 0.5 are shown)

Measures	Factor 1	Factor 2
Stopping time	0.92	
Search time	0.89	
Fixation time	0.79	
Number of fixations	0.78	
Percentage overlap	0.75	
Interfixation distance		-0.86
Percentage area covered		0.84
Percentage detected		0.74
Percentage variance	47	34

ance of 47%) loaded heavily on the speed measures, fixation time, number of fixations and percentage overlap, and appears to represent a 'search speed' factor. Factor 2, with a total percentage variance of 34% represents a 'search accuracy' factor, loaded in accuracy, percentage area covered and (negatively) on interfixation distance. These correlation results were to be expected from models of search, and from empirical evidence. The mean search time for either random or systematic search is (from Drury et al., 1980):

$$\bar{t} = t_o n = \frac{t_o A}{a},$$

where,

- A = total search area,
- a = area of fixation (here the viewer area),
- n = number of fixations, and
- t_o = mean fixation time.

The number of fixations required for coverage is proportional to (A), inversely proportional to (a) and inversely proportional to the degree of fixation overlap. Finally, the stopping time should be (and is) proportional to the mean search time (Tsao et al., 1979). Thus search models predict the intercorrelations of all measures in Factor 1. Similarly, for Factor 2, the accuracy depends upon the coverage of the area, itself a function of the interfixation distance. Finally, there were no significant correlations between the pretest measures (EFT, MFFT) and any process or performance measures.

Table 4
Percentage changes in performance measures and process measures after training

	Control group	Statistical process group	Graphical group	Performance group
<i>Process measures</i>				
% Area covered	2	–7	2	1
Number of fixations	7	–54	–31	–28
Interfixation distance	–5	15	–7	0
Percentage overlap	6	–55	–51	–18
<i>Performance measures</i>				
Search time	0	–26	–15	–37
Stopping time	–5	–50	–35	–51
Percentage detected	–1	–8	2	–1

Note: Negative values indicate decrease.

3. Discussion

With respect to the significant effects observed across the four groups, some patterns in the results were more important than others. Table 4 summarizes the changes between Trial 1 and Trial 2 for each group.

From Table 4, it is clear that no changes in performance measures occurred for the Control group, in contrast to the three feedback groups. All three feedback groups improved in speed, although the Statistical Process group had a reduction in accuracy. Given that these performance changes occurred, the strategy changes underlying each performance seen in the Table 4 can be considered.

The Control group showed significant strategy changes, which should have increased the overall accuracy of search. Thus, more viewer fixations were made, at shorter interfixation distances, resulting in more overlap. However, the (non-significant) decrease in viewer fixation time prevented this increased coverage from either worsening search speed or increasing search accuracy. The reason for the strategy changes and the resulting poor performance of the Control group can only be hypothesized. To compensate for the lack of feedback information and its potential for influencing the effectiveness and efficiency of their search strategy, subjects scanned the area for a longer time before terminating the

search process. This resulted in higher stopping times and a larger number of fixations.

If the Control group had the smallest performance improvement, the Performance group had the largest, reducing search time by about 30% and stopping time by about 50%, without decreasing search accuracy. Clearly, as others have found (Parkes and Rennocks, 1971; Czaja and Drury, 1981; Micalizzi and Goldberg, 1989), outcome or performance feedback enhances performance. This performance gain resulted from a more efficient strategy, with less fixations and less overlap between fixations. In the inspection of printed circuit boards, Schoonard et al. (1973) found that better inspectors used fewer fixations, while Bhatanger (1987), quoted in Drury (1991), found that microcircuit inspectors who were more experienced had a strategy with less, and less overlapped, field of view fixations. Thus the low overlap strategy appears to be one associated with high performance inspectors in industry. In fact, this decrease in overlap and viewer fixations was significant for all three feedback groups, showing that all the subjects in the three feedback groups took on this aspect associated with better inspectors.

The Statistical Process group received task information in a statistical form, with all of the process measures given in digital form. The speed of the Statistical Process group increased, but at the expense of accuracy. Like the other feedback groups, the Statistical Process group reduced the number of viewer fixations and percentage overlap, suggesting a more efficient search. However, they also reduced their coverage of the search field both by increasing the interfixation distance and reducing the area covered. Apparently the insights into their own policies and strategies (Balzer et al., 1989) caused them to concentrate on improving these strategies as all process measures moved towards greater efficiency. Improvements in efficiency in all measures simultaneously will not improve performance as these measures trade off against each other. All that the Statistical Process group achieved by concentrating on the process measures was to move along their speed/accuracy operating characteristic (SAOC) curve in the direction of greater speed (almost doubled) and less accuracy (almost three times the errors). It appears that the insight did indeed change the strategy as predicted but, in the absence of performance data,

these changes did not benefit performance. Without performance feedback, subjects were not completely able to realize the impact of strategy changes on inspection performance, thus showing degradation in detection performance with improved search efficiency.

Finally, the Graphical (visual) group, who received graphical feedback on the scanning pattern, showed perhaps the best combined response in performance and strategy. Their speed improved without affecting their error rate. They evolved a combination of efficiency improvements (fewer viewer fixations, shorter interfixation distances, less overlap) which did not compromise coverage, unlike the Statistical Process group. Their insight was working for them even in the absence of performance feedback. The visual display was designed to be highly intuitive, representing sequence, overlap and coverage by using colored tiles to illustrate successive viewer fixations. Thus, subjects could see uncovered areas directly, without having to translate a number. They also could see their search sequence, a piece of information lacking from the statistical process feedback display. Strictly, such additional information is not needed to evaluate strategy. However, a subject must eventually operationalize changes in strategy into a search sequence. Hence, providing a direct visual representation of this sequence was apparently beneficial, compared with providing the numbers which summarized this sequence. The bridge between cognitive data and action was apparently provided by the visual feedback display. In the review of cognitive feedback studies, Balzer et al. (1989) state: "the pattern of findings from the studies reviewed, however, suggests that TI (Task Information) has been the primary contributing element, whereas CI (Cognitive Information) has been of little value in enhancing performance" (p. 423).

If performance is indeed the final goal of feedback (and it would be difficult to convince the training community otherwise), then the TI provided by the successive-marker visual display was the task information our subjects needed to succeed. TI can be any or all of cue information, criterion information or cue/criterion relationships. All three were provided in a compatible manner by the visual display, but were not as obvious in the statistical display used by the Statistical Process group. Clearly

the form of the display of cognitive feedback requires careful selection if the benefits are to be gained in terms of better process understanding.

4. Conclusions

In this inspection task, simulating an airframe inspection, feedback of three different forms was superior to a control condition of practice without feedback. However, the effect of each form of feedback was different. performance feedback enhanced performance, while cognitive feedback (both statistical and graphical) had a larger effect on the process measures (used in evaluating the visual search strategy). Cognitive feedback using a graphical display of the scanning pattern also enhanced visual inspection performance. These results indicate that if we are to use cognitive feedback to develop a better understanding of the process (search strategy) and hence the performance, it is critical to select the appropriate form of display of cognitive feedback.

Acknowledgements

This research was supported by a contract to SUNY at Buffalo from the Federal Aviation Administration's Office of Aviation Medicine (contract monitor: Dr. W. Shepherd) via Galaxy Scientific Corporation.

References

- Arani, T., Karwan, M.H. and Drury, C.G., 1984. A variable memory model of visual search. *Human Factors*, 26(6): 631–639.
- Baker, C.A., Morris, D.F. and Steedman, W.C., 1960. Target recognition on complex displays. *Human Factors*, 2: 51–61.
- Balzer, W.K., Doherty, M.E. and O'Conner, R.O., 1989. Effects of cognitive feedback in performance. *Psychological Bulletin*, 106: 410–433.
- Bhatanger, V.W., 1987. The effects of experience on search strategy, performance and stress during microscope inspection of real IC chips. Unpublished Ph.D. dissertation, University at Buffalo.
- Bloomfield, J.R., 1975. Theoretical approaches to visual search.

- In: C.G. Drury and J.G. Fox (Eds.), *Human Reliability in Quality Control*. Taylor and Francis, London.
- Brogan, D., Gale, A. and Carr, K. (Eds.), 1993. *Visual search 2*. In: *Proc. 2nd International Conference on Visual Search*. Taylor and Francis, London.
- Brogan, D. (Ed.), 1990. *Visual search*. In: *Proc. first International Conference on Visual Search*. Taylor and Francis, London.
- Chaney, F.B. and Teel, K.S., 1967. Improving inspector performance through training and visual aids. *Journal of Applied Psychology*, 51: 311–315.
- Cockrell, J.T. and Sadacca, R., 1971. Training individual image interpreters using team consensus feedback. Technical Research Report 1171, U.S. Army Behavior and Systems Research Laboratory. In: D.D. Embrey, 1979. *Approaches to Training for Industrial Inspection*. *Applied Ergonomics*, 10: 139–144.
- Czaja, S.J. and Drury, C.G., 1981. Training programs for inspection. *Human Factors*, 23(4): 473–484.
- Deane, D.H., Hammond, K.R. and Summers, D.A., 1972. Acquisition and application of knowledge in complex inference tasks. *Journal of Experimental Psychology*, 92: 20–26.
- Doherty, M.E. and Balzer, W.K., 1988. Cognitive feedback. In: B. Brehmer and C.R.B. Joyce (Eds.), *Human Judgement: The SJT approach*. North-Holland, Amsterdam.
- Drury, C.G., 1990. Visual search in industrial inspection. In: D. Brogan (Ed.), *Visual Search*. Taylor and Francis: London, pp. 263–276.
- Drury, C.G., 1991. Predicting search performance for multiple targets. *Human Factors*, 22(6): 707–718.
- Drury, C.G. and Addison, J.L., 1973. An industrial study of the effects of feedback and fault density on inspection performance. *Ergonomics*, 16(2): 159–169.
- Drury, C.G. and Clement, M.R., 1978. The effects of area, density and number of background characters in visual search. *Human Factors*, 20: 597–602.
- Drury, C.G. and Kleiner, B.M., 1993. Design and evaluation of an inspection training programme. *Applied Ergonomics*, 24(2): 75–82.
- Drury, C.G., Morawski, T.B. and Karwan, M.H., 1980. Inspection performance. In: G. Salvendy (Ed.), *Handbook of Industrial Engineering*. John Wiley, New York, ch. 88, pp. 2282–2314.
- Drury, C.G., Prabhu, P. and Gramopadhye, A.K., 1990. Task analysis of aircraft inspection activities: Methods and findings. In: *Proc. Human Factors Society 34th Annual Meeting*, Orlando, Florida.
- Drury, C.G. and Wang, M.J., 1986. Are research results in inspection task specific? In: *Proc. Human Factors Society 30th Annual Meeting*, pp. 476–480.
- Embrey, D.E., 1975. Training the inspector's sensitivity and response strategy. In: C.G. Drury and J.G. Fox (Eds.), *Human Reliability in Quality Control*. Taylor and Francis, London, pp. 11–16.
- Gallwey, T.J., 1986. Selection tests for visual inspection on a multiple fault type task. *Ergonomics*, 28(5): 595–606.
- Gallwey, T.J. and Drury, C.G., 1986. Task complexity in visual inspection. *Human Factors*, 28, 5.
- Gramopadhye, A.K., Drury, C.G. and Sharit J., 1993. Training for decision making in aircraft inspection. In: *Proc. Human Factors and Ergonomics Society 37th Annual Meeting*, pp. 1267–1271.
- Gramopadhye, A.K., Drury, C.G. and Prabhu, P., in press. Training for visual inspection. *Applied Ergonomics*.
- Greening, C.P., 1976. Mathematical modeling of air-to-ground target acquisition. *Human Factors*, 18: 111–148.
- Hammond, K.R., 1971. Computer graphics as an aid to learning. *Science*, 172: 903–908.
- Hammond, K.R., Summers, D.A. and Deane, D.H., 1973. Negative effects of outcome-feedback in multiple-cue probability learning. *Organizational Behavior and Human Performance*, 9(1): 30–34.
- Human Factors Society, Inc., 1979. *Human Factors*, 21(3).
- Karwan, M.H., Morawski, T.B. and Drury, C.G., 1995. Optimum speed of visual inspection using a systematic search strategy. *IEE Transactions*, 27: 291–299.
- Krendel, E.S. and Wodinski, J., 1960. Visual search in unstructured fields. In: A. Morris and E.P. Horne (Eds.), *Visual Search Techniques*. National Academy of Sciences, Washington, DC, pp. 151–169.
- Lamar, E.S., 1960. Operational background and physical considerations relative to visual search problems. In: A. Morris and E.P. Horne (Eds.), *Visual Search Techniques*. National Academy of Sciences, Washington, DC.
- Lindell, M.K., 1976. Cognitive and outcome feedback in multiple cue probability learning tasks. *Journal of Experimental Psychology: Human Learning and Memory*, 2: 739–745.
- Megaw, E.D. and Richardson, J., 1979. Eye movements and industrial inspection. *Applied Ergonomics*, 10(3): 145–154.
- Micalizzi, J. and Goldberg, J., 1989. Knowledge of results in visual inspection decisions: Sensitivity or criterion effect. *International Journal of Industrial Ergonomics*, 4: 225–235.
- Morris, A. and Horne, E.P., 1960. *Visual Search Techniques*. National Academy of Sciences, Washington, DC.
- Morawski, T., Drury, C.G. and Karwan, M.H., 1980. Predicting search performance for multiple targets. *Human Factors*, 22(6): 707–718.
- National Academy of Sciences, 1973. *Visual Search*. Authors, Washington, DC.
- Parkes, K.R., and Rennocks, J., 1971. The effect of briefing on target acquisition performance. Technical Report 260, Loughborough University of Technology, Loughborough.
- Parker, G.C. and Perry, G., 1972. Lighting for glassware production, Parts II and III. BGIRA Tech Note, 157, Sheffield, UK.
- Schoonard, J.W., Gould, J.D. and Miller, L.A., 1973. Studies of visual inspection. *Ergonomics*, 16(4): 365–379.
- Schwabish, S.D. and Drury, C.G., 1984. The influence of reflective-impulsive cognitive style on visual inspection. *Human Factors*, 26(6): 641–647.
- Shepherd, W.T., Johnson, W.B., Drury, C.G., Taylor, J.C. and Berminger, D., 1991. *Human Factors in Aviation Maintenance, Phase I: Progress Report*, DOT/FAA/AM-92/16, Galaxy Scientific Corporation. National Tech Information Service, Springfield, VA.

- Todd, F.J. and Hammond, K.R., 1965. Differential effects in two multiple-cue probability learning tasks. *Behavioral Science*, 10: 429–435.
- Tsao Y-C., Drury, C.G. and Morawski, B.T., 1979. Human performance in sampling inspection. *Human Factors*, 21(1): 99–105.
- Weiner, E.L., 1975. Individual and group differences in inspection. In: C.G. Drury and J.G. Fox (Eds.), *Human Reliability in Quality Control*. Taylor and Francis, London.
- Williams, L.G., 1966. The effects of target specifications on objects fixated during visual search. *Perception and Psychophysics*, 27: 355–360.